

**AB-18 - Paper****Towards a Tri-Service Model of Selection for the Australian Defence Force**

**J. E. Greig and S.H. Bongers**  
**Defence Force Psychology Organisation**  
**Defence Personnel Executive**

The Defence Reform Program has seen the integration of all personnel management functions into one program, the Defence Personnel Executive. With the intent of achieving greater commonality, efficiency and operational capability, the Defence Personnel Executive is working towards a number of Tri-Service initiatives including the development of a Tri-Service platform for recruitment and selection to the Australian Defence Force. This paper summarises the development of this platform including, the identification of a Tri-Service model of applicant screening, interviewing and testing. Specifically the paper reports on the proposed application of two computer-based tests drawn from the British Army Recruit Battery (BARB) as part of an up-front screening battery. Results on BARB obtained from a sample of applicants for commission and enlistment to the RAAF are reported, and for a small sample of RAAF trainees and cadets validity coefficients yielded by the BARB tests and composite scores are compared.

The Defence Personnel Executive (DPE) was established to achieve efficiencies by integrating the personnel functions of the Royal Australian Navy, the Australian Army and the Royal Australian Air Force. As part of that reorganisation, the three single-Service psychology organisations were amalgamated into a Defence Force Psychology Organisation (DFPO).

Since the amalgamation, the DFPO has been working to achieve more cost efficient selection procedures that will prove effective in providing the Australian Defence Force (ADF) with the best available personnel. The new selection procedures include two-stage testing at Australian Defence Force Recruiting Units (ADFRUs). Under this model, all applicants for entry to the ADF will be administered the same general ability tests. Applicants for occupations in which there are inherent requirements for specific abilities or previous learning will proceed to second-stage testing with relevant aptitude and/or achievement tests.

Against this background, the Director of Defence Force Recruiting (DDFR) requested the introduction of a short pre-screening test at Defence Force Career Reference Centres (DFCRCs) across the Country. If pre-screening could be implemented successfully, processing loads at the seven ADFRUs would be reduced and DDFR would be able to lower the significant costs associated with transporting applicants from regional centres to the larger recruiting units.

DDFR's request was timely because our second report of the Australian trial of the British Army Recruit Battery (BARB) had pointed to the utility and potential of the battery. That report presented data supporting the hypothesis that the battery measures intelligence as that term is understood in the psychometric tradition (Bongers & Greig,

19990423019

1997). One indication of the BARB's construct validity was the finding that an exploratory factor analysis of the six tests comprising the battery identified two factors that are interpretable using intelligence related constructs.

Relevant to the feasibility of DDFR's request for a short screening battery was the associated finding that the factor loadings of Test SA and Test ND indicated that those variables could be treated as surrogate representatives of the first and second factors. Also relevant were findings from subsequent factor analyses, which showed that SA and ND also loaded with two established intelligence tests that we were using as markers. In turn, a composite formed by combining the scores on SA and ND was found to have substantial correlations with the marker tests. Taken together, these findings suggested that a ten-minute battery comprising the two tests might provide valid estimates from testing.

Consistent with these indications, an analysis of the data from the total sample of 3407 applicants for enlistment or commissioning in the Royal Australian Air Force showed that this new composite variable was near normally distributed and as gender-fair as the General Trainability Index (GTI), the composite variable computed from six BARB tests. As would be expected of a measure of intelligence, the composite computed from the two BARB tests measured across a wide range of general ability and yielded statistically significant differences between the mean score from applicants for enlistment and the mean score from applicants for commissioning. Importantly, as well as being statistically significant, a useful effect size (0.78 SD) was associated with the difference between those mean scores.

The consistent findings suggested the potential usefulness of the short battery for, as succinctly stated by Kline (1991), 'Intelligence tests correlate positively with almost all abilities and with a wide variety of real-life criteria.' Given DDFR's requirement for pre-screening at DFCRCs, we have changed the name of the composite from C1 to the Australian Defence Force Index (ADFI). Its particular advantages for pre-screening include short administration times that will further reduce costs by facilitating the scheduling of applicants for testing, and the availability of norms computed from a large sample of applicants for enlistment or commissioning in the Royal Australian Air Force. To these should be added the advantages of invariant administration and accurate scoring that are associated with computer-delivered tests, and the unique advantages offered by the BARB system itself.

Dann, Tapsfield and Collis (1997) explicate the theory, research and development of the BARB computer-delivered test system. This system is innovative because the program generates its test items in the form of elementary cognitive tasks (ECTs) that require only functional levels of literacy. Scores on the BARB tests depend on cognitive processes, not on high levels of educational attainment (Tapsfield & Wright, 1993). The item-generative algorithms produce what essentially are parallel forms at each test administration, thereby facilitating the task of providing applicants with shorter test-retest intervals.

While offering these advantages, however, the reliability and predictive validity of the ADFI must be scrutinised and evaluated against the options of pre-screening with one or

more of the selection tests in current use. The first sets of criterion data for the BARB trial have been collected and while those sets comprise small to very small numbers an initial evaluation of the ADFI is now possible.

This study was aimed at achieving two objectives. First, to confirm the two factor structure of the BARB tests initially reported in our Part 2 study of the Australian trial of the British Army Recruit Battery (Bongers & Greig, 1997). Secondly, to compute and compare validity coefficients yielded by the GTI, by the ADFI, by the two tests used to compute the ADFI, and by the selection tests in current use.

## **Method**

### **Subjects**

Subjects for the first study were the 3407 applicants for enlistment or commissioning in the Royal Australian Air Force who were scheduled for selection testing at ADFRUs between 1 July 1996 and 30 June 1997. The enlistment group included 967 males and 427 females aged between 16 and 35 years. Those who applied for commissioning included 1519 males and 494 females aged between 16 and 43 years. Small sub-sets of the total applicant group were the subjects for the validation studies.

### **Design**

As regards our first objective, the independent variables were two measurement models applied to six of the seven tests that comprise the British Army Test Battery (BARB) Version AC. As the seventh test (PJ) has been dropped from the battery, it was not included in this study. Dependent variables were the scores on each test yielded by the 3407 subjects.

In relation to our second objective, the independent variables were index and test scores from the BARB, the RAAF Commission Test Battery (COMITB), and the RAAF Groundstaff Test Battery (GTB). Dependent variables were scores on four military training courses, the average academic mark awarded by the University of New South Wales to RAAF first-year cadets at the Australian Defence Force Academy (ADFA), and results for those cadets on the military subjects Defence Studies and Military Law.

### **Apparatus**

The BARB tests were administered at ergonomically designed test stations, each furnished with a Pentium 75 microcomputer equipped with 8Mb of RAM and a 685 Mb hard disk drive. Test responses were entered by way of a Microtouch 15-inch touch screen interface. A copy of the BARB software was installed on every hard disk drive, and computers were linked to a Hewlett Packard HP5/100 server for the purpose of collecting and printing each applicant's scores. All computers were connected by means of a twisted-pair Ethernet using RJ-45 connectors. The operating system for the BARB program was MSDOS 6.22, with Windows NT 3.51 installed on the server.

### **Materials**

Materials included Version AC of the BARB software, which included algorithms to generate the ability tests and routines to score responses, transform raw scores to T-scores and calculate the GTI. The composite AGTI was computed from corrected raw scores on the BARB tests SA and ND using the procedure described by Tapsfield (1995).

The selection tests administered to applicants were the authorised batteries used to determine test eligibility for entry to the Royal Australian Air Force. Although different *specialist* batteries were administered, all applicants for enlistment were administered three tests used to calculate the RAAF General Ability Index (G Index). These are: WA (word knowledge), MX (arithmetic) and C (clerical abilities). All applicants for commissioning were administered Test B42, a general ability test published by ACER but restricted for use by the Australian Defence Force.

## Procedure

Two weeks before the day of testing, applicants were notified that a computer delivered test battery would be administered in addition to the standard paper and pencil tests used in the RAAF selection process. A BARB booklet was included, and applicants were advised to read the booklet and complete the items before attending on the scheduled test day.

The selection batteries were administered using RAAF Psychology Service standard operating procedures, including timed breaks at stages of testing. After completing the relevant selection batteries applicants were provided with a 15-minute break before the BARB administration. Applicants were informed that the BARB tests were part of a process aimed at introducing computer administered tests, and that they would not be 'screened-out' for poor performance on the battery. The applicants were advised to perform to the best of their ability because their results on the computer administered tests would be considered along with other possible compensating factors should their results on the pencil and paper tests be below the required standard.

Data from the trial was analysed using SYSTAT Version 7.01 and Amos Version 3.6 software packages.

## Results and Discussion

### Factor Structure

The first investigation was focussed on confirming the two-factor structure of the BARB tests initially reported by Bongers and Greig (1997). As that first factor analysis used T-scores computed with British Army norms, all data used in the confirmatory study were restandardised on the Australian sample. As a check, the exploratory analysis was repeated using this new data set.

Table 1 presents the pattern matrix from the replicated maximum likelihood factor analysis using direct oblimin rotation with gamma set at zero. This analysis used scores

from the 3407 applicants for either enlistment or commissioning who were administered BARB for the first time between 1 July 1996 and 30 June 1997.

Table 1  
Rotated pattern matrix for BARB test scores after  
a maximum likelihood factor analysis

BARB Test	Loadings	
	Factor 1	Factor 2
SA	0.8246	-0.0675
T2	0.7444	0.0356
LC	0.5930	0.1112
ND	-0.0726	0.8804
RF	0.1092	0.4981
A2	0.2939	0.3718

Notes. 1 The two-factor solution explains 51.31 percent of the total variance.  
 2 Factor 1 explains 58.28 percent and Factor 2 explains 41.72 percent of common variance.  
 3 The correlation between the two oblique factors is 0.7682

The notes under Table 1 show that the two-factor solution explains 51 percent of the total variance, and that the two highly correlated factors explain respectively 58 percent and 42 percent of the common variance. As expected, the loadings lead to the same interpretable two factor solution reported and discussed in the earlier study (Bongers & Greig, 1997).

Although the methodology of maximum likelihood factor analysis yielded an interpretable two factor solution, we note that British studies using principal components analysis have consistently reported single factor solutions with moderate to high component loadings (Tapsfield, 1993; Tapsfield, 1995; Kitson & Elshaw, 1996).

In view of the different outcomes from the two exploratory approaches, we decided to evaluate the alternative solutions with a confirmatory procedure. To this end, we specified both an unrestricted model with one factor and a restricted model comprising two correlated factors. Graphical representations of the two models are at Appendix A.

Table 2 presents some measures of fit associated with the alternative models. The measures of fit shown in the table include those implicitly recommended by Browne and Mels (1992), with the exception that ECVI has been replaced by MECVI because maximum likelihood is the default estimation method of the Amos program.

Table 2

Statistics showing some measures of fit associated with two models for the same set of BARB data. Interpretative indications stated.

Measures of Fit	Indications of Fit		Specified Models	
	Saturated	Notes	One Factor	Two Factor
CMIN	0.000	0.000	269.035	72.847
P			.000	.000
F MIN	0.000	0.000	.079	.021
RMSEA <sup>1</sup>		See Note 1	.092	.049
RMSEA 90% CI (Lo)		Confidence	.083	.039
RMSEA 90% CI (Hi)		Intervals	.102	.059
PCLOSE <sup>2</sup>		See Note 2	.000	.554
NCP	0.000		260.035	64.847
FO	0.000		.076	.019
FO 90% CI (Lo)	0.000		.062	.012
FO 90% CI (Hi)	0.000		.093	.028
GFI	1.000		.973	.993
NFI	1.000		.959	.989
CFI <sup>3</sup>		See Note 3	.961	.990
MECVI	0.012		.086	.029

Notes 1. Browne and Cudeck (1983) are of the opinion that an RMS error of approximation of about 0.08 or less would indicate a reasonable error of approximation. They suggest that an RMSEA of .05 or less indicates a close fit

2. PCLOSE tests the null hypothesis that the population RMSEA is no greater than .05. It gives a test of 'close fit' in contradistinction to P, which gives a test of exact fit (Arbuckle, 1997, at page 559).

3. The Comparative Fit Index (Bentler, 1990). CFI values close to 1 indicate a very good fit (Arbuckle, 1997, at page 560).

CMIN is distributed as chi-square and P is the 'p value' for a test of the hypothesis that the model being evaluated fits perfectly in the population. While the P statistic associated with each model provides evidence against the null hypothesis, this evidence is not conclusive because:

It is generally acknowledged that most models are useful approximations that do not fit perfectly in the population. In other words, the null hypothesis of perfect fit is not credible to begin with and will in the end be accepted only if the sample is not allowed to get too big (Arbuckle, 1997 at page 554).

Because of this problem, many statistics less sensitive to sample size have been proposed to assist the process of evaluating the fit of a model. A number of these are reported in Table 2 along with statistics referenced to a 'saturated' or extreme model that is so general it would provide a perfect fit to any set of data. Where a saturated value is not stated, notes provide suggestions to assist interpretation of the relevant observed statistic. Inspection of the measures presented in Table 2 will show that the two-factor model provides the better overall fit on every comparison.

Given that, when sample sizes are very large, the chi-square test will detect small differences between the data-sourced covariances and those that are implied by the particular model, the statistic none-the-less serves the process of evaluation by providing a method for testing which of two alternative models fits the same set of data better. This

chi-square difference test involves a direct comparison of the competing models, the new chi-square statistic and its degrees of freedom being obtained by subtracting the respective values associated with each model. A resulting non-significant chi-square value would indicate that the overall fits of the two models are comparable.

**Table 3**  
Chi-square test of significance of the difference in fit between a one factor model and a two factor model for the same set of BARB data

Model	Chi Square	d.f.	p
1 factor	269.035	9	
2 factor	72.847	8	
Difference	196.188	1	< 0.0000

The results of a comparison of the two models are presented in Table 3. While the significant chi-square difference value does not mean that the common factor model is *the* model that best fits both the data and the theoretical constructs, it does provide a further reason for our preferring that model to the single factor model. Our conclusion is tentative, however, because it rests on the findings from analyses of our present data set only.

### Validity Coefficients

The second investigation was aimed at identifying validity coefficients by correlating a set of predictors with the available criterion data. However, very small sample sizes are associated with four of the five data sets. To provide a benchmark that would assist interpretation of the validity coefficients from the BARB composites and the two tests identified as surrogates, we included two of the predictors currently used in the RAAF selection process. Those predictors are the G Index and Test B42.

The G Index is a composite that is computed from standardised scores on three tests from the RAAF Groundstaff Test Battery. This composite is used in the process of selecting and classifying applicants for enlistment in the Royal Australian Air Force. Test B42 is a general ability test that is used in the process of selecting applicants for commissioning, either by way of entry to the Australian Defence Force Academy (ADFA) or by way of direct entry officer training. Test B42 is published by the Australian Council for Educational Research and its use is restricted to the Defence Psychology Organisation.

Table 4 presents the correlations between the end of course scores from two RAAF training establishments and the G Index, the GTI, the ADFI, and the two BARB tests that are equally weighted when computing the ADFI.

Table 4

Correlations between predictors and criterion scores

at two RAAF Training Establishments

Test	Criterion	N	Pearson's	Probabilities		Std Error	Confidence Intervals	
			r	Uncorrected	Bonferroni		r	Lower 95% Upper 95%

**1 Recruit Training Unit:**

G Index	EOC score	200	0.2156	0.0022	0.0109	0.0674	0.0834	0.3477
GTI	EOC score	200	0.0952	0.1799	0.8995	0.0701	-0.0421	0.2325
ADFI	EOC score	200	0.1231	0.0825	0.4126	0.0696	-0.0134	0.2596
SA	EOC score	200	0.0753	0.2893	1.0000	0.0703	-0.0625	0.2131
ND	EOC score	200	0.1277	0.0715	0.3577	0.0696	-0.0086	0.2640

**Clerical and Supply Trades School:**

G Index	EOC score	96	0.1420	0.1675	0.8377	0.1000	-0.0540	0.3380
GTI	EOC score	96	0.2154	0.0351	0.1755	0.0973	0.0246	0.4061
ADFI	EOC score	96	0.2335	0.0220	0.1102	0.0965	0.0444	0.4226
SA	EOC score	96	0.2582	0.0111	0.0555	0.0953	0.0715	0.4449
ND	EOC score	96	0.1312	0.2026	1.0000	0.1003	-0.0654	0.3278

Note. Correlations are not corrected for restriction of range.

Considering first the data associated with recruit training, Table 4 shows that only the G Index is significantly correlated with the end of course score. Those data also show that only very small to small coefficients are associated with the BARB predictors. It is possible, given the estimated precision, that the observed correlations with the BARB variables are lower bound estimates, but we are unable to identify any reason to assume that this might be the case.

We note, however, that the observed data could be consistent with the findings of Holroyd, Atherton and Wright (1995a, 1995b). Holroyd *et al* found that BARB scores predicted performance in basic military training, but that the strength of the relationships varied according to the learning demands of the subject matter and the reliability of the particular criterion measure available. In this regard, we note also that the recruit-training course has been described as providing a nurturing academic

environment with low cognitive demands. The course is not difficult academically, and failures are mainly attributable to the physical demands of training. While there is no reason to doubt the reliability of the end of course assessment procedure, it is focussed on the application of knowledge gained during the course and does not call on problem solving ability. As Kline (1993, p.19) points out, the difficulties of establishing predictive validity stem from the problem of finding a clear criterion.

Examining the data for training at the RAAF Clerical and Supply Trades School, we note that, in contradistinction to the pattern of correlations in the data for recruit-training courses, the relationship between the G Index and the available end of course mark is not statistically significant. The data presented in Table 4 show that the strongest correlation was between the criterion and Test SA. Although the Bonferroni adjustments signal a need for caution when considering the statistical probabilities associated with the number of comparisons, the data in the relevant rows of Table 4 show the relative strength of each association between the particular predictor and the criterion score. The sample is very small however, and we note that with an assumed correlation of .26 in the population the power to yield a statistically significant result is only 0.74 percent.

Table 5 presents the correlations between three first-year criteria at the Australian Defence Force Academy (ADFA) and scores on Test B42, the GTI, the ADFI, and the two BARB tests that are equally weighted when computing the ADFI.

Table 5

Correlations between predictors and criterion scores

at the Australian Defence Force Academy.

Test	Criterion	N	r	Pearson's		Probabilities		Std Error	Confidence Intervals	
				Uncorrected	Bonferroni				r	Lower 95%
B42	Academic	104	0.3397	0.0004	0.0063	0.0867	0.1697	0.5097		
GTI	Academic	104	0.2099	0.0325	0.4875	0.0937	0.0261	0.3936		
ADFI	Academic	104	0.2915	0.0027	0.0402	0.0897	0.1157	0.4674		
SA	Academic	104	0.0850	0.3912	1.0000	0.0974	-0.1058	0.2758		
ND	Academic	104	0.3608	0.0002	0.0025	0.0853	0.1936	0.5280		
B42	Military Law	104	0.1790	0.0691	1.0000	0.0949	-0.0071	0.3650		
GTI	Military Law	104	0.2362	0.0158	0.2364	0.0926	0.0548	0.4177		
ADFI	Military Law	104	0.3058	0.0016	0.0239	0.0889	0.1316	0.4800		
SA	Military Law	104	0.3283	0.0007	0.0100	0.0875	0.1568	0.4998		
ND	Military Law	104	0.1863	0.0582	0.8733	0.0947	0.0008	0.3719		
B42	Def. Studies	104	0.1370	0.1654	1.0000	0.0962	-0.0515	0.3256		
GTI	Def. Studies	104	0.1812	0.0656	0.9840	0.0948	-0.0047	0.3671		
ADFI	Def. Studies	104	0.2143	0.0290	0.4343	0.0936	0.0309	0.3976		
SA	Def. Studies	104	0.1594	0.1060	1.0000	0.0956	-0.0279	0.3467		
ND	Def. Studies	104	0.1873	0.0569	0.8535	0.0946	0.0019	0.3728		

Note. Correlations are not corrected for restriction of range.

The correlations presented in Table 5 show the relationship of predictors with three first-year criteria. The criterion labelled 'Academic' is the average academic mark awarded by the University of New South Wales. 'Military Law' and 'Defence Studies' are subjects within the military curriculum. On inspection of the table, the data will show that Test ND yielded the highest correlation with the academic criterion. The current selection test B42 yielded the second highest correlation, followed by the ADFI. Test SA showed the strongest association with marks for Military Law, whereas the ADFI yielded the highest correlation with Defence Studies. The data also show statistically non-significant relationships between Test B42 and both military criteria. In contradistinction, the ADFI yielded the second highest correlation with Military Law and the highest correlation with Defence Studies.

## Conclusions

This study was aimed at achieving two objectives. First, to confirm the two factor structure of the BARB tests initially reported in our Part 2 study of the Australian trial of the British Army Recruit Battery (Bongers & Greig, 1997). Second, to compare the validity coefficients yielded by Test B42, by the GTI, by the ADFI, and by the two tests used to compute the ADFI.

As regards our first objective, a second maximum likelihood factor analysis using the same data set after its re-standardisation with Australian norms replicated an earlier analysis using British Army norms (Bongers & Greig, 1997). Tests SA and ND were again found to yield the largest factor loadings, the size of the loadings suggesting that each test could be thought of as a surrogate measure of its latent variable. Two confirmatory factor analyses provided reasons for preferring a two correlated factor model to an alternative one factor model. While the evidence supporting this preference is clear, that finding does not mean that the particular model specified provides the best fit with both data and theory. However, while much work remains, the structural equation modelling procedures used in the confirmatory analysis provide means to test a wide range of hypotheses in a search for the model that is in best accord with both theoretical constructs and the data.

Turning to the second objective, we note that over the five comparisons involving criterion data, the current selection tests yielded the largest correlation only once. Correlations involving either the AGTI or one of the two tests comprising that composite were larger over the other four comparisons. Again, over the same comparisons, the correlations between all five criterion measures and the AGTI were larger than those between the same criterion measures and the GTI. This observation is very tentative, however, because the low power and precision associated with four of the five comparisons would make nonsense of any claim to find meaning in an ordering of the coefficients in terms of their magnitude. Our samples are too small, and we must wait for more data from the training establishments.

Although we have emphasised the tentative nature of our own observations, they are consistent with some findings from Jacobs and Longmore (1998). In that study, which involved larger sample sizes, the researchers found that Test SA was the best single predictor of performance for seven of 11 courses in Phase II of British Army training. Test ND was the best single predictor for one of the courses, and the second best single predictor for a further five courses.

Our research will continue to focus on gaining a better understanding of the BARB tests; on seeking further evidence of construct validity, and on investigating the validity of both composite scores and individual tests as predictors of training and job performance. With increased sample sizes and broader criterion measures, future studies will aim at identifying the particular predictor-criterion relationships that have the greatest utility value. In the shorter term, research activities will include analysing data from a larger sample of applicants who have been retested in order to estimate standard errors of measurement with greater precision.

## References

ACER (1981). *ACER Higher Tests: ML-MQ (2<sup>nd</sup> edition) and PL-PQ Manual*. Melbourne: Australian Council for Educational Research.

Arbuckle, J. L. (1997). *Amos Users' Guide: Version 3.6*. Chicago, IL: SmallWaters Corporation.

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.

Bongers, S.H. & Greig, J.E. (1996). An Australian trial of the British Army Recruit Battery: Part 2. *Proceedings of the 39<sup>th</sup> Annual Conference of the International Military Testing Association*. Sydney, 14<sup>th</sup> to 16<sup>th</sup> October 1997.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. & Long, J.S. (Eds.). *Testing structural equation models*. Newbury Park, California: Sage, 136-162. Cited in J.L. Arbuckle. (1997). *Amos Users' Guide: Version 3.6*. Chicago, IL: SmallWaters Corporation.

Browne, M.W. & Mels, G. (1992). *RAMONA User's Guide*. The Ohio State University, Columbia, Ohio. Cited in J.L. Arbuckle. (1997). *Amos Users' Guide: Version 3.6*. Chicago, IL: SmallWaters Corporation.

Dann, P., Tapsfield, P. & Collis, J. (1997). The theory, research and development of the British Army Recruit Battery. Human Assessment Laboratory, University of Plymouth.

Holroyd, S.R., Atherton, R.M. & Wright, D.E. (1995a). The criterion-related validity of the British Army Recruit Battery. *Proceedings of the 37<sup>th</sup> Annual Conference of the International Military Testing Association*. Toronto, 16<sup>th</sup> to 19<sup>th</sup> October 1995.

Holroyd, S.R., Atherton, R.M. & Wright, D.E. (1995b). Validation of the British Army Recruit Battery against measures of performance in basic military training. Centre for Human Sciences, Report DRA/CHS/HS3/CR95019/1.0. DRA, Farnborough.

Jacobs, N.R. & Longmore, K. (1988). Validation of the Current Soldier Selection Measures against Phase II Training Performance. Centre for Human Sciences, Report PLSD/CHS/HS3/CR9800085/1.0. Defence Evaluation and Research Agency, Farnborough.

Kitson, N. & Elshaw, C.C. (1996). A Comparison of the British Army Recruit Battery and the RAF Ground Trades Test Battery. Centre for Human Sciences, Report DRA/CHS/HS3/CR96060/1.0. Defence Evaluation and Research Agency, Farnborough.

Kline, P. (1991). *Intelligence: The psychometric view*. London: Routledge.

Kline, P. (1993). *The handbook of psychological testing*. London: Routledge.

Holroyd, S.R., Atherton, R.M. & Wright, D.E. (1995a). The criterion-related validity of the British Army Recruit Battery. *Proceedings of the 37<sup>th</sup> Annual Conference of the International Military Testing Association*. Toronto, 16<sup>th</sup> to 19<sup>th</sup> October 1995.

Holroyd, S.R., Atherton, R.M. & Wright, D.E. (1995b). Validation of the British Army Recruit Battery against measures of performance in basic military training. Centre for Human Sciences, Report DRA/CHS/HS3/CR95019/1.0. DRA, Farnborough.

Jacobs, N.R. & Longmore, K. (1988). Validation of the Current Soldier Selection Measures against Phase II Training Performance. Centre for Human Sciences, Report PLSD/CHS/HS3/CR9800085/1.0. Defence Evaluation and Research Agency, Farnborough.

Kitson, N. & Elshaw, C.C. (1996). A Comparison of the British Army Recruit Battery and the RAF Ground Trades Test Battery. Centre for Human Sciences, Report DRA/CHS/HS3/CR96060/1.0. Defence Evaluation and Research Agency, Farnborough.

Kline, P. (1991). *Intelligence: The psychometric view*. London: Routledge.

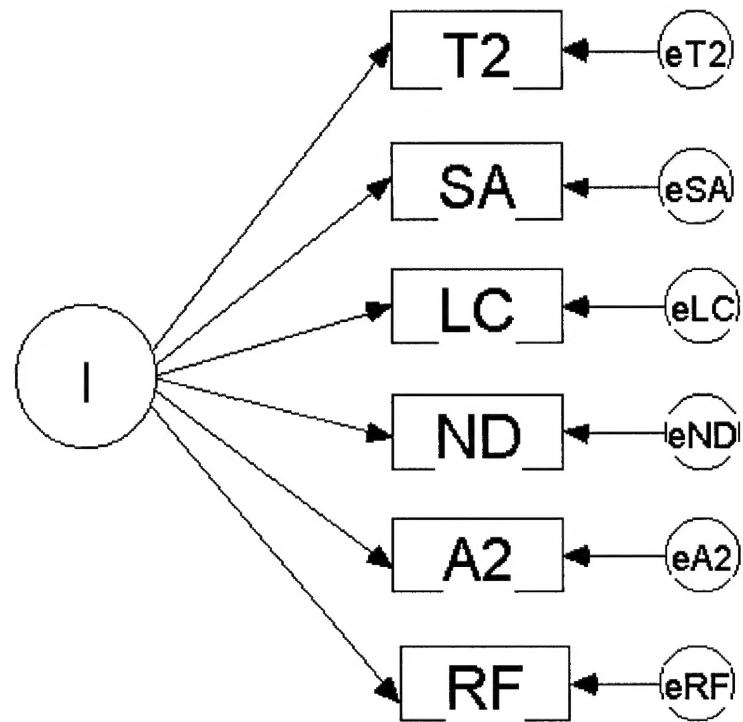
Kline, P. (1993). *The handbook of psychological testing*. London: Routledge.

Tapsfield, P.G.C. (1993). The British Army Recruit Battery: Test-Retest Reliability. HAL Technical Report: 5 – 1993 (APRE). University of Plymouth.

Tapsfield, P.G.C. (1995). The British Army Recruit Battery: 1995 Applicant Norms. HAL Technical Report: 13-1995 (DRACHS). University of Plymouth.

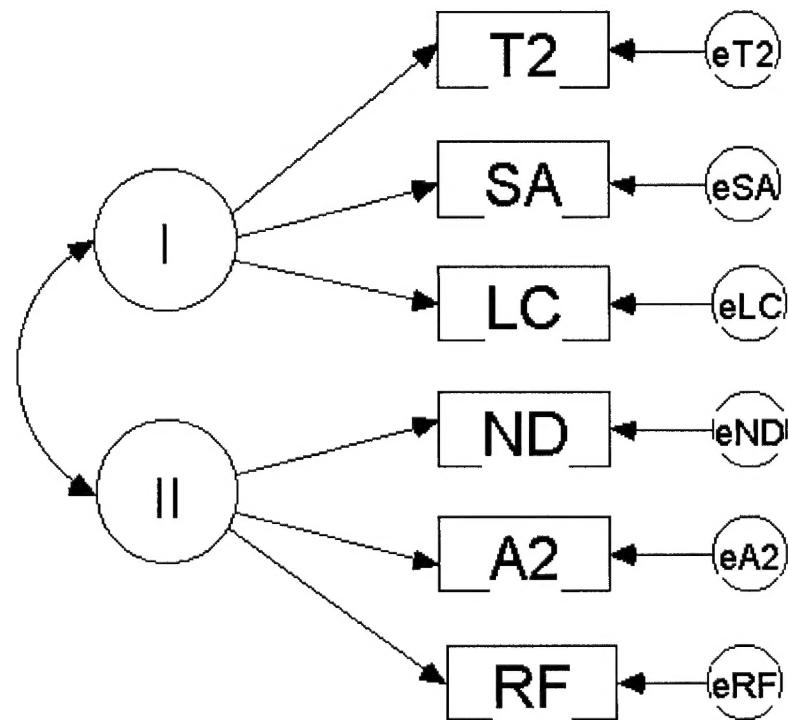
Tapsfield, P.G.C. & Wright, D.E. (1993). A preliminary analysis of summary data arising from the operational use of the British Army Recruit Battery. HAL Technical Report 3-1993 (APRE). University of Plymouth.

## Appendix A



**Figure 1**  
The one factor model

## Appendix A



**Figure 2**

**The two correlated factors model**

## INTERNET DOCUMENT INFORMATION FORM

**A . Report Title:** Towards a Tri-Service Model of Selection for the Australian Defence Force

**B. DATE Report Downloaded From the Internet** 4/21/99

**C. Report's Point of Contact: (Name, Organization, Address, Office Symbol, & Ph #):** Navy Education and Training Professional Development and Technology Center  
Navy Advancement Center Dept  
Dr. Grover Diesel, (850) 452-1815  
6490 Saufley Field Road  
Pensacola, FL 32509-5237

**D. Currently Applicable Classification Level:** Unclassified

**E. Distribution Statement A:** Approved for Public Release

**F. The foregoing information was compiled and provided by:**  
**DTIC-OCA, Initials:** VM\_ **Preparation Date:** 4/22/99\_

The foregoing information should exactly correspond to the Title, Report Number, and the Date on the accompanying report document. If there are mismatches, or other questions, contact the above OCA Representative for resolution.